

## **Generic Metrics for the Estimation of the Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool**

**Mike M. Paglione, Mary Lee Cale, and Dr. Hollis F. Ryan**  
FAA William J. Hughes Technical Center

**An aircraft conflict probe is a strategic tool used by the air traffic controller to predict aircraft flight paths and to identify future conflicts. The FAA has designated the strategic conflict probe as a core function for the future ground based systems required for “Free Flight”. With these systems under development, there is a need for a generic set of metrics to quantify the performance of the conflict probe. This paper presents definitions of the two fundamental measures of the conflict prediction accuracy, missed alert and false alert probabilities. These fundamental probabilities are expanded upon to define a conflict prediction sensitivity measure, referred to as the sharpness metric, and specific examples on how the sharpness measure can be applied are presented.**

### **INTRODUCTION**

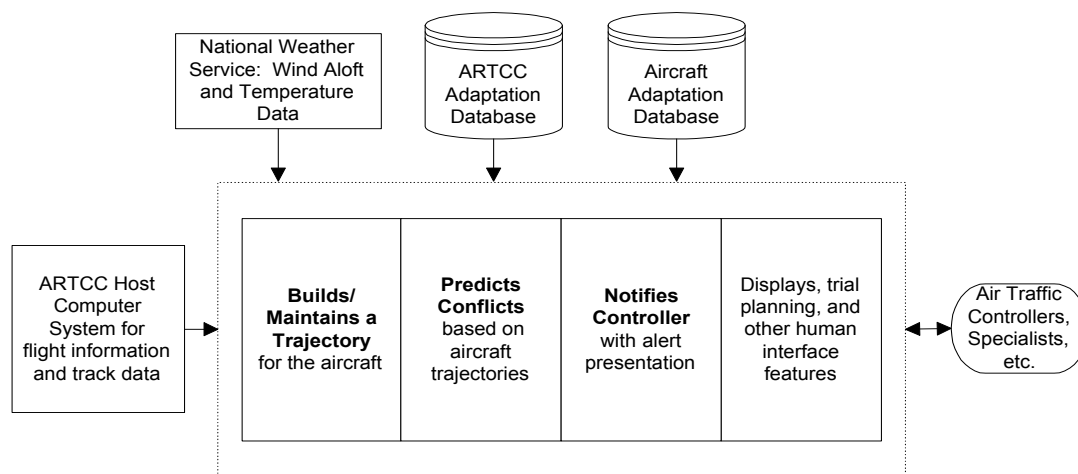
In the United States, the overall system of managing and controlling air traffic is known as the National Airspace System (NAS), which is administered by the Federal Aviation Administration (FAA). Surveillance radar provides aircraft position information to the ground controllers, radio navigation aids provide position information to the pilots, and very high frequency (VHF) radios provide voice communications between the aircraft and the ground. Detailed procedures involving restrictions on routing, speeds, and altitudes are an integral part of the NAS. The restrictions severely reduce the amount of aircraft traffic that NAS can accommodate, particularly when the weather is bad. Free Flight is an air traffic control concept that increases the efficiency of aircraft operations while maintaining safety. It will do this by reducing the restrictions imposed by NAS on the aircraft flights. The goal of free flight is to provide “unrestricted opportunity for all to use the limited airspace in a manner that is efficient, effective, and equitable” (RTCA, 1996).

To achieve the goals of Free Flight, broad categories of advances in ground and airborne automation are required. One of the most important ground based tools currently being developed is a conflict detection tool or conflict probe (CP). A conflict probe is a decision support tool that will provide the air traffic controller with predictions of conflicts (i.e., loss of minimum separation between aircraft) for a parameter time (e.g., 20 to 40 minutes) into the future. At a minimum, a conflict probe predicts the flight path of an aircraft, continuously monitors that flight path from current aircraft position information, and probes for conflicts with other aircraft and incursions into restricted airspace. The tool also assists the controller in resolving the predicted conflicts, and with alternative route planning in response to user requests. In contrast to the current, more tactical methods of air traffic control, a conflict probe supports Free Flight by aiding the controller in the strategic planning of aircraft separation management.

The FAA has sponsored the development of two prototype conflict prediction tools: the User Request Evaluation Tool (URET) developed by MITRE/CAASD and the Center-TRACON Automation System (CTAS) En Route Descent Advisor (E/DA) developed by NASA Ames Research Center. The technical accuracy of these tools is a critical issue to be addressed in planning for Free Flight Phase 1 (FFP1) and the future integration of these tools. NASA Ames and CAASD have created and applied performance metrics for their specific conflict prediction tools (Bilimoria, 1998; Brudnicki et al., 1998). The Traffic Flow Management branch (ACT-250) at the FAA William J. Hughes Technical Center (WJHTC) has defined a generic set of metrics that highlight the performance of any conflict probe (Cale et al., December 1998). Since these metrics are independent of a particular system’s design choices, they provide common measures to evaluate the performance of different systems. Four broad categories of metrics have been defined: trajectory accuracy, conflict prediction accuracy, conflict notification timeliness, and conflict prediction stability. This paper focuses on the fundamental conflict prediction accuracy metrics, missed alert and false alert, and introduces the sharpness metric, a conflict prediction sensitivity measure.

## PROBLEM DESCRIPTION

A conflict probe is responsible for predicting into the future (e.g. 20 minutes) both the path an aircraft will fly, and potential conflicts the aircraft will have with other aircraft or with restricted airspace. As implemented in existing conflict probe prototypes, the aircraft's trajectory and any conflict predictions are based on the flight information and track data from the Air Route Traffic Control Center's (ARTCC) Host Computer System (HCS), weather forecasts from the National Weather Service, and detailed adaptation databases, including aircraft modeling information and system information relating to the airspace and procedures (see Figure 1). The conflict probe uses the flight intent and tracked position information received from the HCS to build and maintain an aircraft trajectory that predicts the flight path of the aircraft. This process can include either monitoring the tracked position compared to the trajectory and rebuilding it when necessary, or rebuilding the trajectory upon receipt of every track report. The common element in maintaining a trajectory in the various prototypes is that the original predicted path or trajectory is changed as more information becomes available, often simply by updating the aircraft's trajectory to match the expanded route determined from the original or controller amended flight plan. By using these trajectories for all the active aircraft, the conflict probe predicts future conflicts with other aircraft and restricted airspace.

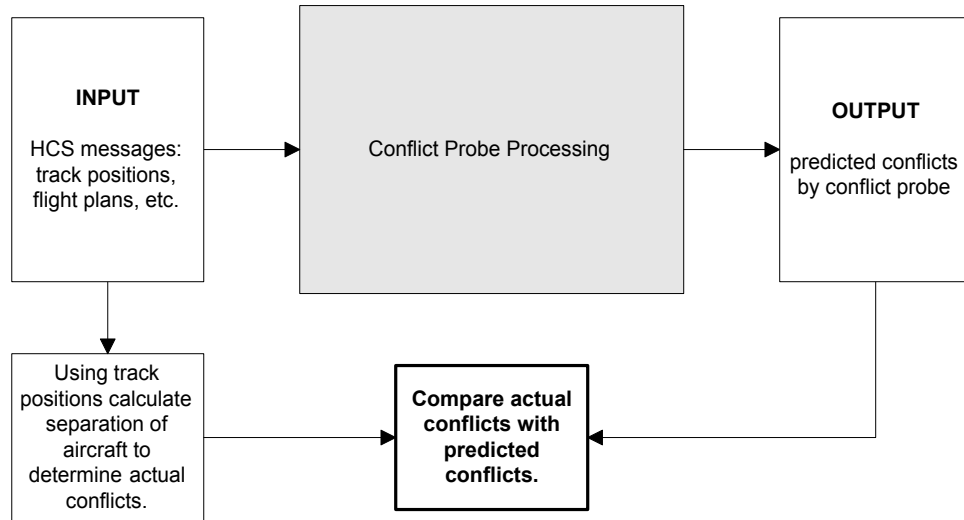


**Figure 1: Components of the Conflict Probe Process**

The accuracy of the conflict probe predictions can be measured in several ways. Four broad categories of metrics have been defined: trajectory accuracy, conflict prediction accuracy, conflict notification timeliness, and conflict prediction stability (Cale et al., December 1998). A conflict probe uses its predicted trajectories to determine future separation violations, i.e., to predict conflicts. Thus, the trajectory accuracy, or the deviation between the predicted trajectory and the actual path of the aircraft, has a direct effect on the accuracy of the conflict prediction. Conflict prediction accuracy is measured by several error probabilities that are used to quantify whether a predicted conflict actually occurred, and whether an actual conflict was predicted. The conflict predictions must not only be accurate in terms of the existence of a separation violation, but the conflict needs to be predicted in a timely manner. Conflict notification timeliness attempts to quantify the amount of lead time the probe provides in the conflict predictions. Finally, the conflict prediction stability metric quantifies the stability of the various predictions made by the conflict probe. For example, a probe can make accurate trajectory or conflict predictions, but if these predictions change too frequently the user will have difficulty in making a choice between them.

The focus of this paper is on the measurement of the accuracy of a conflict probe's predictions of aircraft to aircraft conflicts. This is probably the most operationally significant metric category, since the major purpose of a conflict probe is to support strategic separation management of aircraft. Conflict prediction accuracy quantifies the fundamental error probabilities that are directly related to the probe's central goal: detecting conflicts. Referring to Figure 2, the conflict prediction accuracy metric isolates the conflict probe processing

as a black box. Such an approach is only concerned with the input (i.e. the positions of the aircraft) and the output (i.e. predicted conflicts). A post-processing tool must first determine the actual conflicts using the aircraft position data, and then these conflicts are compared to the predicted conflicts.



**Figure 2: Conflict Prediction Accuracy Process**

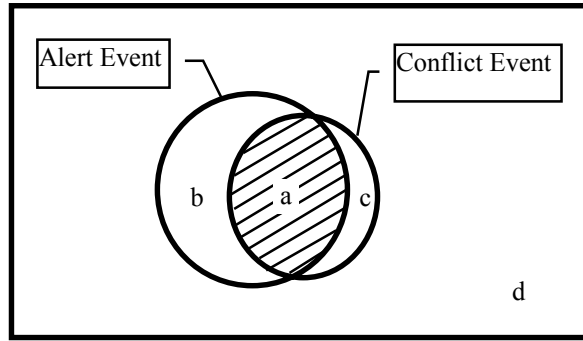
#### DEFINITION OF ERROR EVENTS AND PROBABILITIES

A conflict between two aircraft is defined as the simultaneous loss of separation beyond specified thresholds in both the horizontal and vertical dimensions. An alert is the prediction of a conflict by the conflict probe. The conflict prediction accuracy metrics describe two fundamental events: a conflict occurs and an alert is predicted. These events, which are not mutually exclusive, have four possible outcomes (see Table 1). The conflict accuracy metrics measure these two fundamental error outcomes: missed alert and false alert. For the error outcome defined as a false alert (cell b in Table 1), the conflict probe performs a conflict prediction by presenting an alert without a corresponding conflict. For the error outcome defined as a missed alert (cell c in Table 1), a conflict occurs but the probe does not present a corresponding alert.

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT	CP predicts conflict and it occurs (a -- valid alerts)	CP predicts conflict and it does not occur (b -- false alert)
NO ALERT	CP does not predict conflict and it occurs (c -- missed alert)	CP does not predict conflict and it does not occur (d -- remaining aircraft pairs)

**Table 1: Alert and Conflict Event Table**

The combination of outcomes of these two events is a random process related to the performance of the conflict probe. Therefore, probability will be used to quantify the likelihood or chances of occurrence of the associated random events; that is, the likelihood of incorrect predictions of the conflict probe. Figure 3 depicts the events presented in Table 1 as a Venn diagram. The correct prediction occurs for the two of the four outcomes represented by regions a and d in Figure 3, while the incorrect predictions are represented by the other two regions, b and c.



**Figure 3: Venn Diagram of Alert and Conflict Events**

Now that the two random events have been partitioned into four outcomes, it is important to define the probabilities carefully. It is practically meaningless to quantify the probability of a false alert or missed alert without using conditional probability. For example, the calculation for the direct false alert probability would be the number of predictions that occurred without a corresponding conflict over the total number of aircraft combinations. Since the total number of aircraft combinations is approximately half the squared number of aircraft in the sample set, this probability will be very small. A more meaningful probability calculation would condition this probability on the alert event. In other words, the probability of a false alert given an alert outcome exists is a better measure of the performance, since it should not be as dependent on the size of the sample set. The probability of a false alert and the corresponding sample estimate is expressed in Equation 1. The estimate of the conditional false alert probability is simply the ratio of false alerts over the total number of alerts in the sample.

$$P(FA|A) = \frac{P(C' \cap A)}{P(A)} = \frac{\text{probability of no conflict and an alert takes place}}{\text{probability of an alert}} \approx \frac{\text{number of false alert outcomes from sample}}{\text{number of alerts from sample}} \quad \text{Equation 1}$$

An analogous metric is expressed in Equation 2 for the missed alert probability, which conditions on the conflict event. The missed alert outcome is thus expressed as the probability of a missed alert given a conflict takes place. Similar to the method used for the false alert estimate, the missed alert estimator is the ratio of missed alert outcomes over the total number of conflicts in the sample.

$$P(MA|C) = \frac{P(A' \cap C)}{P(C)} = \frac{\text{probability of no alert and a conflict takes place}}{\text{probability of a conflict}} \approx \frac{\text{number of missed alert outcomes from sample}}{\text{number of conflicts from sample}} \quad \text{Equation 2}$$

### **DISCUSSION OF A NEW METRIC: SHARPNESS**

It is not sufficient to report a missed alert probability without a corresponding false alert probability (and vice versa), since these two fundamental errors are not independent and furthermore are inversely proportional. As a result, it is possible to reduce either one of these probabilities with an increase to the other. For example, to reduce the missed alert probability to very low limits, a conflict probe could expand its conflict separation thresholds to very large values and thus predict conflicts for practically all aircraft pair combinations. Unfortunately, this would not be acceptable to the air traffic controller using the tool, since the false alert probability would probably increase dramatically. Therefore, if the developers design a conflict probe with acceptable limits of both these error probabilities, a balance is implied. This balance between false alert and missed alert probabilities not only must be acceptable, but balance should be robust in terms of the separation between aircraft and other factors. The need to determine the proper trade off between the two

error probabilities, to support both the designer and the FAA sponsors in the evaluation of these tools, prompted research into metrics that model the sensitivity of the conflict predictions.

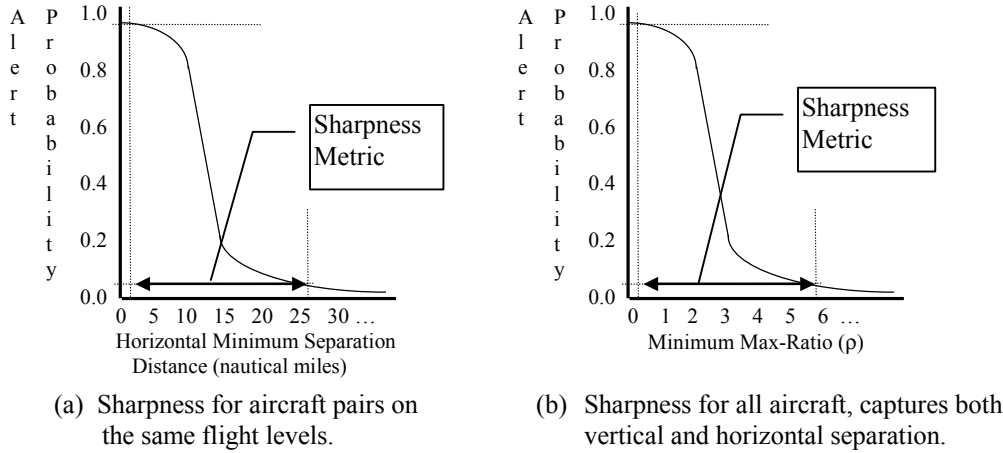
Sensitivity measures are often used with error detection systems by quality control engineers in manufacturing industries. For example, Statistical Quality Control (SQC) Charts are used to detect shifts in a manufacturing process (Montgomery, 1991). These charts detect a change or shift in the process average and are used to minimize the number of defective products. Similar to the conflict prediction of a conflict probe, there are two kinds of errors associated with the detection of a shift in the process mean, referred to as Type I and Type II errors. The Type I error probability refers to the probability of detecting a shift when a shift did not take place; this is analogous to the false alert probability for a conflict probe. The Type II error is the probability of not detecting a shift when the process mean did really make a shift; this is comparable to the missed alert probability. When evaluating a process, quality engineers use sets of curves, referred to as operating characteristic functions, to make a trade off between these two types of measurement errors. The curves are a plot of the probability of not detecting a shift versus an associated actual shift in the process mean. The quality engineer plots these curves for different measurement designs to decide which design best measures the particular process. The design with the curve with the steepest relationship between the probability of missing a shift versus the actual shift magnitude minimizes errors associated with the detection system. This allows the engineer to design an optimal control chart for the particular process under study.

An operating characteristic function or curve, analogous to the function described above, can be defined for a conflict probe. During research and development of the Automated En Route Air Traffic Control (AERA) concept, MITRE generated similar curves, as a function of the minimum horizontal separation of aircraft pairs at the same altitude, and defined a metric referred to as crispness that measured the steepness of the curve (Niedringhaus et al., 1984). This paper expands upon this concept and introduces a new metric called “sharpness”. The sharpness measure includes all aircraft pairs (i.e., not necessarily at the same flight level) and thus provides an indication of the conflict probe’s ability to discriminate between conflicts with different closest approach distances. The sharpness metric and the associated conflict probe performance curve are discussed in more detail in the following paragraphs.

#### ***Conflict Probe Performance Curve***

The conflict probe performance curve (see Figure 4) is formed by plotting the probability that a conflict probe will present an alert for a pair of aircraft as a function of the minimum separation of the two aircraft. The minimum separation of the aircraft reflects how close the aircraft have come to each other along their entire flight. The variable along the horizontal (X) axis is the minimum horizontal separation which would be attained by the two aircraft if they were to fly near each other on the same flight level. The vertical (Y) axis is the probability that, for a specific minimum horizontal separation, the conflict probe will predict a conflict. For example, if two aircraft flight paths will take the aircraft to within 12 nautical miles of each other, the conflict probe represented by the curve in Figure 4 has a 0.6 probability of predicting an alert.

To quantify the steepness of the curve, the sharpness metric is calculated by finding the intersection points of a probability close to one and the performance curve, and a probability close to zero and the performance curve. Specifically, as illustrated in Figure 4a, the distance along the x-axis between these two points defines the sharpness metric for aircraft pairs on the same flight level. To expand the performance curve for all encounters, this paper defines a method to capture both the horizontal and vertical processes so the performance curve can be generalized to include vertical separation of the aircraft as well as horizontal separation.



**Figure 4: Example of Alert Probability Versus Minimum Separation Distance**

#### Generalization of the Conflict Probe Performance Curve to Include Vertical Separation

To generalize the performance curve shown in Figure 4a to consider all aircraft pairs, not just those on the same flight levels, it is necessary to capture both the horizontal and vertical dimensions of separation on the x-axis, since the legal separation of aircraft includes both dimensions. For the horizontal dimension, the standard separation is given in nautical miles (usually 5 nautical miles). For the vertical dimension, the standard separation is presented on a much smaller scale (e.g., 2000 feet for aircraft above 29000 feet). When considering these standard separation values, an aircraft needs 15 times more separation in the horizontal plane than in the vertical. These two dimensions of separation distances are practically independent, but a conflict takes place only if both are violated simultaneously. The sharpness metric has been defined to capture these independent processes in both dimensions into one value that corresponds to the aircraft pair's minimum separation. First, the separation distance in each dimension is normalized, so that both values are on the same scale. This is accomplished by dividing the aircraft to aircraft separation by the standard separation for each time synchronized<sup>1</sup> track position report in each dimension (the standard vertical separation will vary depending on the location of the conflict; i.e., 1000 feet below 29000 feet, and 2000 feet above 29000 feet). These ratios are expressed in Equations 3 and 4.

The ratio of horizontal separation to standard horizontal separation can be expressed as:

$$\lambda_i = \frac{\sqrt{\left(\frac{x_i^a - x_i^b}{\delta_i}\right)^2 + \left(\frac{y_i^a - y_i^b}{\delta_i}\right)^2}}{\delta_i} \quad \text{Equation 3}$$

where

$\delta_i$  = horizontal separation standard for the  $i^{\text{th}}$  synchronized track data point;

$x_i^a$  = x position of the  $i^{\text{th}}$  track point of aircraft a in nautical miles;

$x_i^b$  = x position of the  $i^{\text{th}}$  track point of aircraft b in nautical miles;

and  $y_i^a$ ,  $y_i^b$  are the corresponding y positions

<sup>1</sup> "Synchronized" means that aircraft a is at its  $i^{\text{th}}$  track position at the same time that aircraft b is at its  $i^{\text{th}}$  track position.

The ratio of vertical separation to standard vertical separation can be expressed as:

$$\pi_i = \frac{|z_i^a - z_i^b|}{v_i} \quad \text{Equation 4}$$

where

$v_i$  = vertical separation standard for the  $i^{\text{th}}$  synchronized track data point;  
 $z_i^a$  = altitude position of the  $i^{\text{th}}$  track point of aircraft a in feet;  
 $z_i^b$  = altitude position of the  $i^{\text{th}}$  track point of aircraft b in feet.

Next, the maximum value or max-ratio of  $\lambda$  and  $\pi$  is calculated for each track point and the minimum from all these maximums is determined for each aircraft pair. The following equation expresses the calculation of the minimum of the maximum ratios.

$$\rho = \min_i^k [\max(\lambda_i, \pi_i)] \quad \text{Equation 5}$$

where

$i$  = current  $i^{\text{th}}$  track point ;  
 $k$  = total number of track points.

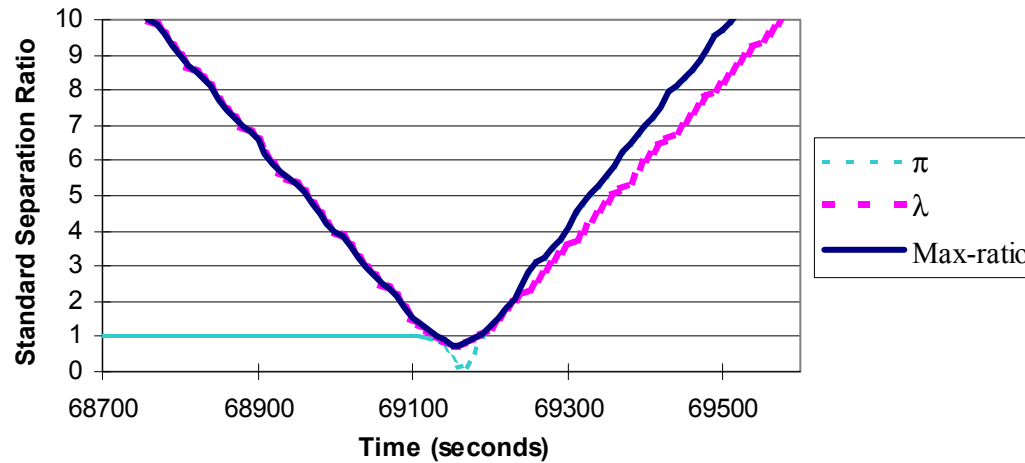
The unitless distance  $\rho$ , referred to as the minimum max-ratio of separation, combines both dimensions of separation and directly corresponds to standard separations. By definition, if  $\rho$  is less than 1, there exists a violation of standard separation; conversely, if  $\rho$  is equal to or greater than 1 there cannot be a violation of standard separation. This new measure,  $\rho$ , is used as the X variable in plotting the performance curve instead of the horizontal separation distance; that is, the probability of an alert being predicted is plotted against the minimum max-ratio  $\rho$ . This generalized curve is illustrated in Figure 4b.

#### **Example of the Application of the Minimum Max-Ratio**

The separation measure defined as the minimum max-ratio, or  $\rho$ , is illustrated by the following example. Consider two aircraft on approaching courses, flying on adjacent flight levels. Aircraft A is flying at 18,000 feet; Aircraft B is at 17,000 feet. The aircraft are initially approaching each other on the same route; as they get closer to each other, Aircraft A starts to diverge from the projected route of Aircraft B. When the aircraft pass each other, going in opposite directions, they have a minimum horizontal separation of approximately 3.5 nautical miles (corresponding to a value of  $\lambda = 3.5/5 = 0.7$ ). This separation is less than the required minimum of 5 nautical miles in en route airspace but, as long as the two aircraft remain at 18,000 feet and 17,000 feet, their vertical separation is sufficient to maintain required separation (the vertical separation of 1000 feet gives  $\pi$  a value of  $1000/1000 = 1.0$ ). However, just before the point of closest horizontal approach, the higher aircraft, Aircraft A, starts to descend. Since the vertical separation is already at the minimum value of 1000 feet, vertical separation is lost immediately and the value of  $\pi$  drops below 1. The simultaneous loss of horizontal and vertical separation causes  $\rho$  to have a value of less than 1.

As the aircraft approach each other, pass, and then diverge from each other, their horizontal separation decreases to the minimum value of 3.5 nautical miles and then increases. The value of  $\lambda$  shown on the graph in Figure 5 starts out at a large value, then decreases to a value of less than 1 before increasing again. As mentioned above, when both aircraft are in level flight, separated in altitude by 1000 feet, the value of  $\pi$  is 1. As shown in Figure 5, when Aircraft A starts to descend, the value of  $\pi$  decreases to values less than 1. As Aircraft A continues its descent, passing through the altitude of Aircraft B, and continuing down, the vertical separation increases to values greater than 1000 feet, and the value of  $\pi$  increase to values above 1. Figure 5

shows how the instantaneous values of  $\lambda$  and  $\pi$ , and their max-ratio, change as the two aircraft fly past each other.



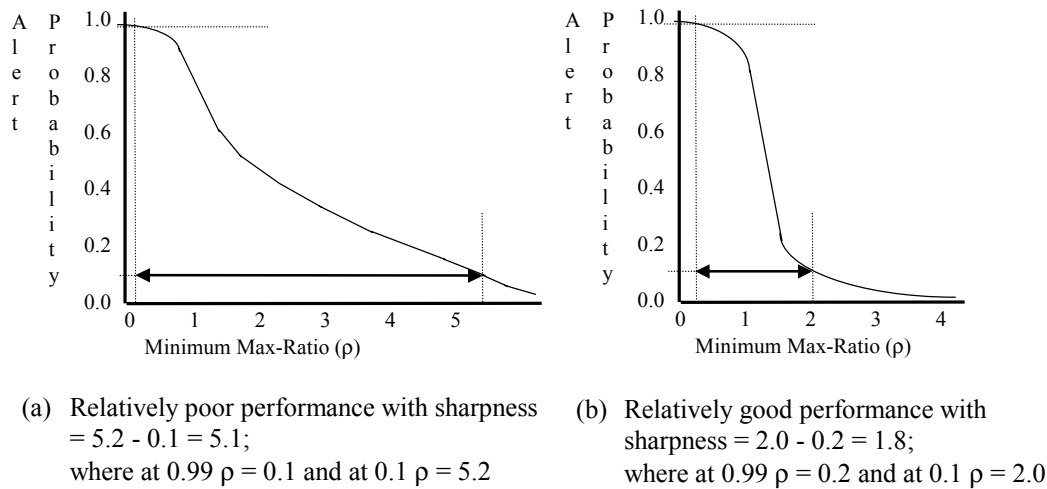
**Figure 5: Plot of Max-Ratio Against Time**

### **Sharpness Metric**

While the missed and false alert probabilities express the absolute conflict prediction errors, the “sureness” or precision of these predictions is represented by the shape of the performance curve. The sharpness metric quantifies the precision of the conflict probe’s conflict prediction and can be determined by measuring the steepness of the conflict probe performance curve. The steeper or more abrupt the incline of the curve, the better the precision of the conflict prediction.

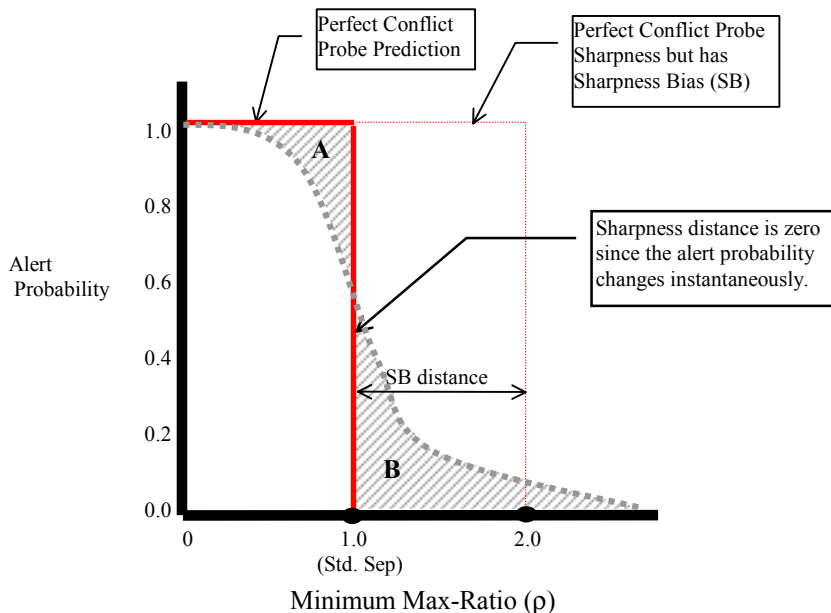
The sharpness metric (S) is the normalized distance measured along the horizontal axis from the point where the probability of an alert is close to one (e.g. 0.99) to the point where the probability of an alert has dropped close to zero (e.g., 0.10). A large value for sharpness, as illustrated in Figure 6a, would indicate a conflict probe that does not work very well. In contrast, a small value for sharpness would be indicative of a steep curve, as shown in Figure 6b, which would mean the conflict probe works well. That is, the smaller the value of the sharpness metric, the better the conflict probe’s predictions. An ideal or perfect conflict probe will have sharpness of zero. The ideal or perfect performance curve, illustrated in Figure 7 as the solid heavy line, is a step function. The probe having this performance curve will always give an alert when the separation is less than the legal minimums and it will never give an alert when the separation is greater than the legal minimums. In other words, the perfect probe would have a probability of one of detecting a conflict with the minimum max-ratio less than one, and a probability of zero of detecting a conflict at a minimum max-ratio of one and greater. This illustrates that the sharpness metric expresses performance of the perfect conflict probe, and shows that the better the performance of the conflict probe under study, the smaller the sharpness distance will be.





**Figure 6: Examples of the sharpness performance curve**

To express again the relationship between the performance curve and the error probabilities of missed and false alerts, the dotted line in Figure 7 exemplifies the performance curve of a more typical conflict probe. The areas labeled A and B represent the missed and false alert probabilities, respectively. The ideal probe has very steep descent eliminating its areas of A and B and thus has no missed and false alert events. The inverse relationship between the two errors is also illustrated in Figure 7. Shifting the curve to the right will decrease the probability of false alerts, but it will certainly increase the probability of missed alerts. Only by increasing the steepness of the curve measured by the sharpness metric will both errors be reduced.

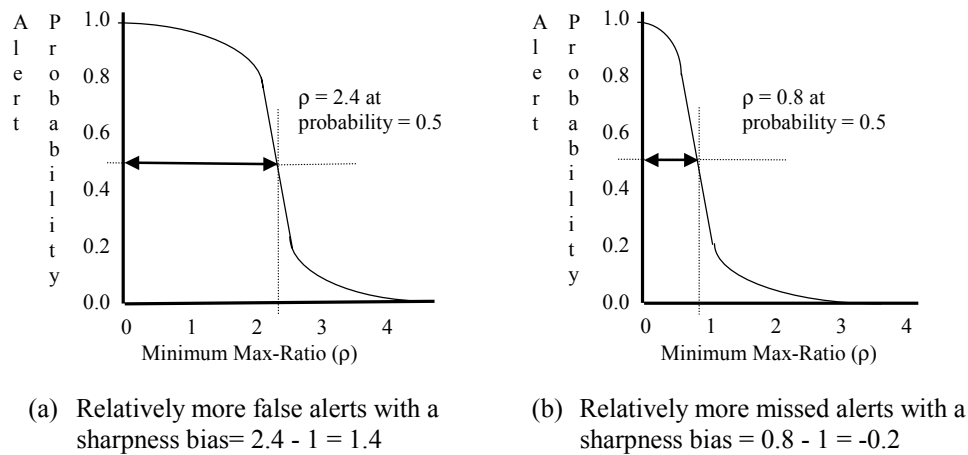


**Figure 7: Ideal or Perfect Conflict Probe Performance**

### Sharpness Bias Metric

To define the ideal or perfect probe, a sharpness of zero is necessary but not sufficient, since the alert probability curve could be offset along the x-axis. A second metric is needed to capture this offset. The offset or separation distance bias, called Sharpness Bias (SB), is defined as the minimum max-ratio (the value of  $\rho$ ) corresponding to an alert probability of 0.5, minus 1 (i.e.,  $SB = \rho - 1$ ). This metric is illustrated in Figure 7 and Figure 8. The perfect probe would have an SB of zero, since the minimum max-ratio at 0.5 probability of alert would be one (see Figure 7). As shown by the dashed line step function in Figure 7, a probe could be very precise as measured by a sharpness of zero but if it has sharpness bias it would still produce errors. In other words, a conflict probe with very precise predictions of encounters may not be absent of error.

The SB represents the conflict probe's built in tolerances used for conflict prediction. A positive value for SB would indicate the performance curve has a bias greater than the defined separation for conflicts and would tend to favor false alerts over missed alerts (refer to Figure 8a). Conversely, a negative value for SB indicates the performance curve has a bias less than the defined separation for conflicts and would tend to favor missed alerts over false alerts (see Figure 8b). That is, SB indicates the extent to which the detection behavior of the conflict probe is conservative relative to the nominal separation standards. It provides an indication of the extent to which the conflict probe designers have moved the performance curve in order to balance missed and false alert probabilities.



**Figure 8: Examples of Sharpness Bias Performance Curves**

### SHARPNESS AND SHARPNESS BIAS CASE STUDY

In February 1998, a simulation study was completed at the FAA William J. Hughes Technical Center to determine the conflict prediction accuracy of the URET prototype Delivery 3 in single center operation (Cale et al., April 1998). The URET prototype is in "daily use" at the Indianapolis and Memphis ARTCCs and is the basis for the FFP1 conflict probe (known as URET Core Capabilities Limited Deployment (CCLD)). ACT-250's approach in accomplishing this study was to develop an Indianapolis simulation capability at the WJHTC. The simulation approach was chosen, rather than using actual "real world" data, because one would expect there to be no conflicts in the actual data since any potential conflict would have been resolved by the controller prior to its occurrence. Utilizing a Host Computer System that resides at the Technical Center, the simulation activity extracted real flight plans from Indianapolis Center System Analysis Recording (SAR) tapes and used a high fidelity aircraft simulator to model the flights without controllers separating the aircraft. This process introduced a number of aircraft to aircraft conflicts. Use of a simulation allowed the conflicts to be modeled at any minimum separation desired as well as the emulation of situations that could not be observed or completely controlled in the real world.

Nine simulations were conducted, each around 5 hours in duration with approximately 400 to 500 simulated aircraft. These nine simulations were analyzed twice with two different definitions of aircraft to aircraft conflicts. URET predicts two levels of alerts for aircraft to aircraft conflicts: red alerts for the violation of standard separation between trajectory center lines, and yellow alerts for violation of standard separation between conformance boxes around the trajectory (the conformance boxes are nominally 2.5 nautical miles laterally, 1.5 nautical miles longitudinally and 300 feet vertically for aircraft in straight and level flight; they are expanded in the appropriate dimension when an aircraft is turning or climbing/descending, and for non-RNAV equipped aircraft). This study did not differentiate between color coding and captured all the alerts. The first analysis, referred to as Analysis A, used basic standard radar separation for en route airspace as defined in FAA Order 7110.65, 4-5-1.a/b and 5-5-3.b.1 (i.e., five nautical miles in the horizontal dimension, and 1000 feet at or below FL290/2000 feet above FL290 in the vertical dimension). This represents the true standard separated conflict situation. The second analysis, referred to as Analysis B, expanded the separation distance in the horizontal dimension to ten nautical miles, which more closely models the encounter distances URET uses in its predictions of yellow alerts for aircraft to aircraft conflicts.

In determining the alert probability for URET, only the first alert for a particular conflict situation was counted, since it represented the earliest alert notification for the particular conflict. Alerts that occurred very close to the initial point of conflict (e.g. less than 5 minutes) were included in this count and were flagged as late "valid" alerts for future study. The impact of these late valid alerts was minimal, since they occurred very rarely.

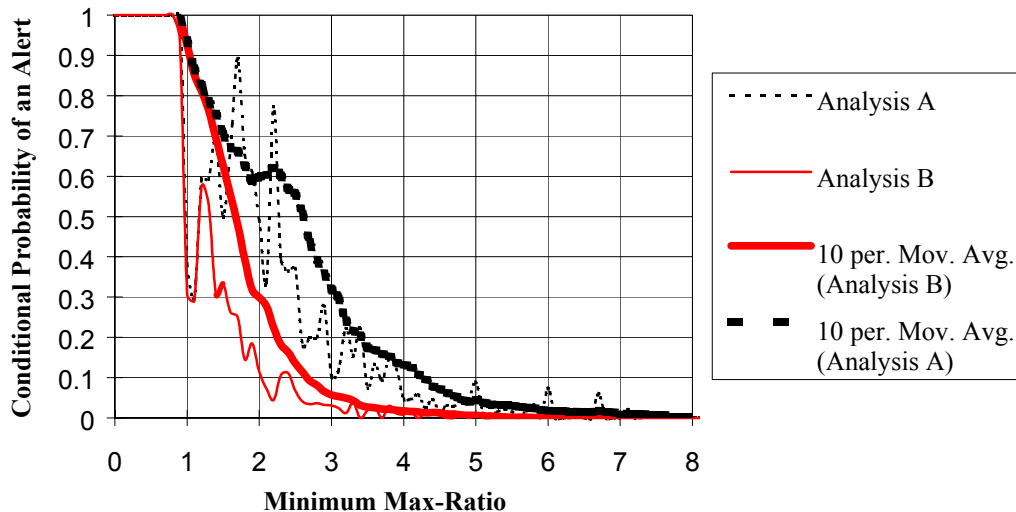
To calculate sharpness (S) and sharpness bias (SB), the HCS track reports of the aircraft from the nine simulations were compared to the predicted aircraft trajectories and the actual horizontal and vertical separations were used to calculate the minimum max-ratio of all aircraft combinations. Next, the URET alerts were matched with the minimum max-ratio (i.e.  $\rho$ ) values calculated from the track reports. The probability of an alert is a conditional probability of an alert with a certain range of  $\rho$ , given the existence of conflicting aircraft pairs with this range of  $\rho$ . To estimate this conditional probability, the number of alerts presented with the specified range of  $\rho$  are calculated from HCS track reports and are divided by the total number of aircraft pairs with the same range of  $\rho$ .

For this study, an interval of 0.1  $\rho$  was used for calculating the performance curve. To estimate the curve, a histogram is formed with probabilities for each 0.1 interval. To calculate sharpness, the difference along the x-axis is calculated between two  $\rho$  values chosen from translating the points from a probability close to 1 to a probability close to 0. The parameters chosen for this study were 0.99 and 0.10. These values were chosen to capture the distance sharpness is designed to measure, namely the sensitivity of conflict predictions to the true separation of aircraft. The upper threshold is 0.01 less than 1.0 and the lower threshold of 0.10 is ten times that probability distance from 0 probability. The chosen thresholds also emphasize the greater significance of a missed alert compared to a false alert and reduce the sensitivity of the sharpness measurement to random fluctuations of the performance curve for large values of  $\rho$  (i.e.  $\rho > 5$ ). Further study with controller reactions to the conflict probe in simulations and field trials will need to be performed to validate these thresholds.

Interpolation is used to translate the probability thresholds to the appropriate  $\rho$  value on the x-axis. In Figure 9, two conditional alert probability versus minimum max-ratio plots are presented for one of the nine simulation runs. The first set of curves (thin dashed line) represent the alerts for Analysis A where the conflicts were based on standard separation distances. The next set of curves (thin solid line) represent the alerts for Analysis B where the conflicts were based on expanded separation distances (i.e. 10 nautical miles horizontal separation). These curves are the plot of the actual probability estimates for each interval. Since the actual probability estimates are hard to compare because of the sampling noise, a set of smoothed moving average plots are also presented in Figure 9. As expected, since an aircraft pair will be in violation of a 10 nautical mile separation standard earlier than a 5 nautical mile standard, the sharpness for Analysis B is significantly smaller than Analysis A in this simulation run. The change in sharpness between these two

analysis curves is around  $1.5 \rho$ . The sharpness bias is also smaller for Analysis B, since the alert probability curve that crosses the 0.5 probability is around one  $\rho$  value smaller than Analysis A.

By calculating the sharpness for all nine simulations, the average sharpness and sharpness bias for the standard separation conflicts of Analysis A are approximately 3 and 1.1, respectively. The average sharpness and sharpness bias for the expanded separation conflicts of Analysis B are approximately 1.4 and 0.2, respectively. The results illustrate that URET is over twice as sensitive to expanded 10 nautical mile conflicts compared to standard five nautical mile conflicts. The results are not surprising, since URET predicts conflicts at greater than standard separation distances, nominally at 10 nautical miles or greater.



**Figure 9: Plot of Conditional Probability of an Alert vs. Minimum Max-Ratio for Both Analysis A and B**

## SUMMARY

The Traffic Flow Management Branch (ACT-250) at the FAA WJHTC has defined four broad categories of generic metrics to quantify the performance of a conflict probe: trajectory accuracy, conflict prediction accuracy, conflict notification timeliness, and conflict prediction stability (Cale et al., December 1998). The focus of this paper is on conflict prediction accuracy, with an emphasis on a new measure called sharpness. In estimating the conflict prediction accuracy of a conflict probe, two fundamental errors are present: missed alerts and false alerts. A conflict probe is designed to meet acceptable limits of both these errors and a balance between the two is implied. However, probabilities of these errors do not describe the sensitivity of the conflict predictions as a function of actual separation. To accomplish this, an aggregate metric, called sharpness, was developed which measures the precision and sensitivity of the conflict predictions. In a sense, sharpness measures the spread of the predictions as a function of the actual aircraft separation distances determined from HCS track data and indicates the ability of the conflict probe to discriminate between conflicts with different minimum separations. Similar metrics were used in the past to measure the sensitivity for aircraft pairs on the same flight level; the sharpness metric expands on this by capturing all aircraft pairs in a given scenario with the minimum max-ratio separation measure.

There are many potential applications of the sharpness metric. This paper presents an example of its application to the URET conflict probe, considering two different definitions of conflicts (i.e. Analysis A's standard separation and Analysis B's expanded separation). As expected, the results suggest that URET makes conflict predictions with both smaller sharpness and sharpness bias for expanded conflicts as opposed to conflicts defined at the separation standards. The sharpness metrics are recommended as a relative

measure when comparing the performance of different conflict probes, different site adaptations, or different parameters used in a conflict probe. For example, different sites with various field adaptations and traffic mixes will certainly create different demands on the performance of a conflict probe. The sharpness and sharpness bias could provide the sensitivity measure on how the performance varies from site to site. It could also provide a useful gauge to the designer/developer on adjustments to the conflict probe parameters to optimize the performance of the conflict predictions, and a useful set of metrics for the FAA to use to compare various approaches in making these predictions. The most effective use of this metric requires further detailed designed experiments to be conducted to determine what factors have a statistically significant effect on sharpness and thus the conflict prediction precision.

## **ACRONYMS**

ARTCC	Air Route Traffic Control Center
ATC	Air Traffic Control
CP	Conflict Probe
CTAS	Center-TRACON Automation System
FAA	Federal Aviation Administration
FFP1	Free Flight Phase 1
HCS	Host Computer System
NAS	National Airspace System
S	Sharpness Metric
SB	Sharpness Bias
URET	User Request Evaluation Tool
WJHTC	William J. Hughes Technical Center

## **REFERENCES**

- Bilimoria, K. (1998), "A Methodology for the Performance Evaluation of a Conflict Probe", AIAA-98-4238, AIAA Guidance, Navigation and Control Conference, Boston, MA, August.
- Brudnicki, D., Arthur, W., Lindsay, K. (1998), *URET Scenario-based Functional Performance Requirements Document*, MTR98W0000044, MITRE/CAASD, McLean, VA.
- Cale, M., Paglione, M., Ryan, Dr. H., Timoteo, D., Oaks, R. (April 1998), *URET Conflict Prediction Accuracy Report*, DOT/FAA/CT-TN98/8, FAA William J. Hughes Technical Center/ACT-250, Atlantic City, New Jersey.
- Cale, M., Paglione, M., Ryan, Dr. H., Timoteo, D., Oaks, R., Summerill, S. (December 1998), "Application of Generic Metrics to Assess the Accuracy of Strategic Conflict Probes", 2<sup>nd</sup> USA/Europe ATM R&D Seminar, Orlando, FL, December.
- Devore, J.L. (1987), *Probability and Statistics for Engineering and the Sciences, Second Edition*, Brooks/Cole Publishing Company, California.
- Paglione, M., Cale, M., Ryan, H. (1997), *Generic Metrics and Statistics to Estimate the Conflict Prediction Accuracy of Conflict Probe Tools* (Draft), FAA William J. Hughes Technical Center/ACT-250, Atlantic City, New Jersey.
- Kelton, D., Law, A. (1991), *Simulation Modeling And Analysis, Second Edition*, McGraw-Hill, Incorporated, New York.
- Montgomery, D. C. (1991), *Introduction to Statistical Quality Control, Second Edition*, John Wiley and Sons, Incorporated, New York.

Niedringhaus, W., Rockman, M., Hannet, H., Hutchinson, J., Stanley, M. (1984), *Performance Analysis of AERA Algorithms: Phase II, Volume 3*, MTR-108, MITRE Corporation, McLean, VA.

RTCA Inc. (1996), *Free Flight Action Plan*, Washington, DC.

## **BIOGRAPHIES**

Mike Paglione is a Software Engineer with the FAA at the William J. Hughes Technical Center where he is the Conflict Probe Algorithm Assessment group lead in the Traffic Flow Management branch. His focus is on developing metrics and simulations to assess the accuracy of conflict probe algorithms. He received his BS and MS degrees in Industrial and Systems Engineering from Rutgers University.

Dr. Ryan is a Senior Systems Engineer with the Signal Corporation involved in the testing and evaluation of aircraft trajectory models. He holds a B.A.Sc. in Engineering Physics from the University of Toronto, and MS and PhD degrees in Electrical Engineering from the University of Illinois.

Mary Lee Cale is a Computer Scientist at the William J. Hughes Technical Center where she is the FAA lead of the En Route Area Work Team of the FAA/NASA Interagency ATM Integrated Product Team. She holds a BS in Mathematics from Oklahoma State University and an MS in Computer Science from Drexel University.